DOCUMENT RESUME

ED 226 064 TM 830 111

AUTHOR Stricker, Lawrence J.
TITLE Test Disclosure and Retest Performance on the Scholastic Aptitude Test.
INSTITUTION College Entrance Examination Board, New York, N.Y.; Educational Testing Service, Princeton, N.J.
REPORT NO CB-R-82-7; ETS-RR-82-48
PUB DATE 82
NOTE 16p.; Some tables may be marginally legible due to small print.
AVAILABLE FROM College Board Publications, Box 886, New York, NY 10101 ($4.00).
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *College Entrance Examinations; High Schools; *Performance Factors; Scores; *State Legislation; *Test Coaching; Testing Problems; Testing Programs; *Test Wiseness
IDENTIFIERS Retesting; *Scholastic Aptitude Test; *Test Disclosure

ABSTRACT

The aim of this study was to evaluate the effect of disclosing a Scholastic Aptitude Test (SAT) form on the retest performance of examinees who initially had been tested with the disclosed form and subsequently retested with a different form. Retest performance was compared for three random samples of examinees who had been tested with the SAT as high school juniors in the May 1981 administration in New York and then retested with it in the October 1981 administration: the standard set of disclosed material for the May SAT was sent to two experimental groups, along with either a noncommittal or an encouraging letter intended to vary their motivation to use the material; nothing was sent to the control group. The three groups were generally similar in the level, stability, and concurrent validity of their October scores, indicating that access to the disclosed material had no appreciable effects on retest performance. The absence of differences for the two experimental groups and for subgroups within them that would have been most apt to use the material suggests that use of the material had no discernible effect either. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED226064

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

S. W. Gardner

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

College Board Report

No. 82-7

Test Disclosure and Retest Performance on the Scholastic Aptitude Test

Lawrence J. Stricker

Test Disclosure and Retest Performance on the Scholastic Aptitude Test

Lawrence J. Stricker

Educational Testing Service

College Board Report No. 82-7

ETS RR No. 82-48

College Entrance Examination Board, New York, 1982

The author wishes to thank Donald L. Alderman, John A. Centra, Philip K. Oltman, Donald E. Powers, Donald A. Rock, and Warren W. Willingham for advising about the research design and statistical analysis; Donald Schiariti for supervising the assembling and mailing of the disclosed material; Peter E. Smith for arranging the retrieval of the data; Patricia W. Cox for statistical calculating; Norma A. Norris for computer programming; and Gordon A. Hale, Donald E. Powers, and Gretchen W. Rigol for critically reviewing a draft of this report.

Researchers are encouraged to express freely their professional judgment. Therefore, points of view or opinions stated in College Board Reports do not necessarily represent official College Board position or policy.

The College Board is a nonprofit membership organization that provides tests and other educational services for students, schools, and colleges. The membership is composed of more than 2,500 colleges, schools, school systems, and education associations. Representatives of the members serve on the Board of Trustees and advisory councils and committees that consider the programs of the College Board and participate in the determination of its policies and activities.

Additional copies of this report may be obtained from College Board Publications, Box 886, New York, New York 10101. The price is $4.

Copyright © 1982 by College Entrance Examination Board.
All rights reserved.
Printed in the United States of America.

CONTENTS

Abstract 1

Introduction 1

Method 2

- Procedure 2
- Measures 2
- Statistical Analyses 2

Results and Discussion 3

- Background Characteristics 3
- Score Level 5
- Retest Reliability 7
- Concurrent Validity 7

Conclusions 9

References 10

Tables

1. Sex of the Samples 4
2. Ethnicity of the Samples 4
3. Father's Education of the Samples 4
4. Mother's Education of the Samples 4
5. Parents' Income of the Samples 4
6. Financial Need of the Samples 4
7. High School Type of the Samples 5
8. High School Programs of the Samples 5
9. High School Rank of the Samples 5
10. High School GPA of the Samples 5
11. Educational Aspiration of the Samples 5
12. Means and Standard Deviations (SD) for Initial, Retest, and Covariance-Adjusted Retest Scores 6
13. Summary of One-way Analyses of Variance and Analyses of Covariance of Initial and Retest Scores 6

14. Correlations Between Initial and Retest Scores 8

15. Correlations of Initial and Retest Scores with Criteria 8

ABSTRACT

The aim of this study was to evaluate the effect of disclosing a Scholastic Aptitude Test (SAT) form on the retest performance of examinees who initially had been tested with the disclosed form and subsequently retested with a different form. Retest performance was compared for three random samples of examinees who had been tested with the SAT as high school juniors in the May 1981 administration in New York and then retested with it in the October 1981 administration: the standard set of disclosed material for the May SAT was sent to two experimental groups, along with either a noncommittal or an encouraging letter intended to vary their motivation to use the material; nothing was sent to the control group. The three groups were generally similar in the level, stability, and concurrent validity of their October scores, indicating that access to the disclosed material had no appreciable effects on retest performance. The absence of differences for the two experimental groups and for subgroups within them that would have been most apt to use the material suggests that use of the material had no discernible effect either.

INTRODUCTION

Public disclosure of the content of admissions tests, originally mandated by legislation in New York and now a nationwide policy of many admissions testing programs (Brown 1980; "Test-Takers" 1981), has potentially important consequences for the performance of examinees. Although there has been a great deal of speculation about this subject (see Brown 1980; Strenio 1979), data are scarce. It is well established, though, that very few examinees seek the disclosed material for most admissions tests, with the striking exception of the Law School Admissions Test (see Linn 1982). The only information on the effects of disclosure on test performance comes from a study of the specific recall of disclosed material (Hale *et al.* 1980). This experiment, in a classroom setting, found that the examinees achieved substantially elevated scores on special forms of the Test of English as a Foreign Language, which consisted of questions already disclosed to the students. These effects occurred whether or not the questions were discussed in class, but the extent of the effects depended on the size of the pool of disclosed items: the examinees who had been given many more disclosed questions than what subsequently appeared on the special forms of the test obtained lower scores. The broader impact of disclosure in the more realistic situation in which examinees repeat a test, after taking an entirely different form of the test, which is subsequently disclosed, and then receiving its questions and their answers, has thus far not been investigated. This issue is of considerable practical importance in view of the substantial proportion of examinees who repeat admissions tests (for example, Donlon and Angoff 1971) and of the weight that admissions officers attached to retest scores (for example, Educational Testing Service 1981a).

In principle, access to the disclosed material in such circumstances, in common with retaking a test, receiving test coaching, and using test orientation materials, such as guidebooks and practice tests (for example, Educational Testing Service 1979, 1980), has the potential for increasing examinees' familiarity with a test's instructions and content, reducing their anxiety about it, and providing an opportunity for them to drill on specific types of questions (Anastasi 1981; Messick 1980). Accordingly, insofar as disclosure has any impact over and above these other influences, subsequent retest scores may be affected in two distinct ways. First, the scores may be elevated. All the possible effects of disclosure that were just mentioned should contribute to score improvement. It is noteworthy that retaking a test and test coaching both produce some score gains (see Anastasi 1981; Messick 1980).

Second, the retest scores may not measure the same thing as the initial scores. Greater familiarity with the nature of a test and reduced anxiety should lead to a more veridical assessment of ability whereas intensive drilling should produce a distorted appraisal (Messick 1980; Messick 1981). Hence, validity may increase or decrease, depending upon the relative importance of these two kinds of influences. Retest reliability may be lowered in any event, for both influences would reduce the correspondence between initial and retest scores. However, the sparse data that are available on these points, based on initial scores and ordinary retest scores on the Scholastic Aptitude Test (SAT), suggest that the effects may be relatively small, at least for test familiarization and anxiety reduction. Although these influences should make the two scores diverge, the scores had similar validity in predicting college grades (Olsen and Schrader 1959), and the retest reliability of the scores is extremely high, approximately .9 (Donlon and Angoff 1971).

A related matter is that these effects on retest scores, rather than being uniform, may vary systematically with the examinees' characteristics. These include variables such as unfamiliarity and anxiety that may lead to poor test performance and be alterable by exposure to the disclosed material, as well as other variables such as motivation and ability that may determine access to the material and effective use of it. Data are lacking on this issue.

The primary aim of this study was to evaluate the effect of disclosing an SAT form on the retest performance of examinees who had been initially tested with the disclosed form and subsequently retested with a different form. More specifically, the goals were to determine whether or not receiving the disclosed test affected (a) the level of retest scores, (b) their retest reliability, and (c) their concurrent validity against a criterion of high school grades. A secondary goal was to explore whether or not the effects depended on the examinees' characteristics—

demographic variables, as well as those that may affect performance and be alterable by exposure to the disclosed material, and those that may determine access and use of the material.

METHOD

Procedure

Three random samples, each consisting of 2,500 examinees, were drawn from those taking the SAT (Form 1Z) in the May 2, 1981, administration in New York. The samples were limited to examinees with the following characteristics, as determined from the registration form and other records about the administration:

1. Junior in high school.
2. Resident of New York.
3. Registered on time for a Saturday administration.
4. SAT-verbal and SAT-mathematical scores were both available for the administration.

Two of the samples, the not encouraged and encouraged experimental groups, were sent the standard set of disclosed material (the operational items on the test, a copy of the examinees' answer sheet, scoring instructions, and a key) that is routinely provided to those who request it. The mailing took place at approximately the same time (June 26-30) that the disclosed material was sent to the first of the May examinees who had requested it.

The material for the two experimental groups was accompanied by a letter from the College Board intended to vary motivation to use the disclosed material. The letters for the two groups differed. The letter for the not encouraged group consisted of a single paragraph:

> Although you may not have requested them, I am sending the questions and answers, as well as a copy of your own answer sheet, for those parts on the May SAT that counted toward your scores on the test. The College Board is sending these materials, on an experimental basis, to a cross-section of all students who took the test.

The letter to the encouraged group contained the same paragraph and an additional one:

> In the event that you plan to take the SAT again, you may find these materials useful in preparing for the test. They should help you to become more familiar with the instructions and the kinds of questions used, and may make it possible for you to take the test with greater confidence.

Nothing was mailed to the third sample, the control group.

Subsequently, the examinees in each of the three groups who were retested with the SAT (Form 1Y) in the October 10 and 11, 1981, administration were identified after excluding three examinees in the not encouraged group and four in the encouraged group to whom the disclosed material could not be delivered. The number of examinees with either SAT-verbal or SAT-mathematical scores available for the administration were 1,248 for the control group, 1,229 for the not encouraged group, and 1,272 for the encouraged group. Of these examinees, 87 in the control group, 59 in the not encouraged group, and 62 in the encouraged group had requested the disclosed material for the May administration.

Measures

The measures used in the statistical analysis appear below. They were obtained from test records for the May and October administrations, and the Student Descriptive Questionnaire completed for the May administration.

SAT Scores

1. May SAT-verbal converted score
2. May SAT-mathematical converted score
3. October SAT-verbal converted score
4. October SAT-mathematical converted score

Background Variables

1. Sex
2. Ethnicity
3. Father's education
4. Mother's education
5. Parents' income
6. Financial need
7. High school type
8. High school program
9. High school rank[1]
10. High school grade point average (GPA)[2]
11. Educational aspiration

Criteria

1. High school rank
2. High school GPA

Statistical Analyses

All statistical analyses were limited to the three samples of examinees who retook the SAT in the October administration and had verbal or mathematical scores available for the administration. Because of missing data for SAT scores and background variables, the sample sizes fluctuated for the analyses; each analysis was based on all the available data.

1. High school rank was quantified as follows: highest tenth = 5, second tenth = 15, second fifth = 30, middle fifth = 50, fourth fifth = 70, and lowest fifth = 90.

2. High school GPA is the mean GPA in six subjects (English, mathematics, foreign languages, biological sciences, and social studies) weighted by the number of years of study per subject. In computing this mean, A = 4, B = 3, C = 2, D = 1, and F = 0; 1 year = 1, 2 = 2, 3 = 3, 4 = 4, and 5 or more = 5. (See Educational Testing Service 1981b.)

SAT-verbal and SAT-mathematical scores were analyzed separately throughout. Parallel analyses were carried out for the May and October SAT scores in order to determine the extent to which the October results were attributable to differences in the composition of the samples rather than to differences in their retest performance. Although the original samples were comparable by virtue of being randomly drawn, self-selection could have produced differences in the fractions of these samples that were retested. Because any sample differences in the May results for the retested examinees are presumably due to such variations in sample composition, it can likewise be assumed that similar differences in the October results for these examinees have the same cause.

Background Characteristics. Sample differences in background variables were assessed by χ^2 tests for discrete variables and by one-way analyses of variance for continuous variables.

Score Level. Sample differences in May and October SAT means were assessed by one-way analyses of variance. Differences in the October SAT means were also appraised by one-way analyses of covariance, controlling for the pertinent May SAT scores (for example, May SAT-verbal score was the covariate in the analysis of October SAT-verbal score). Interactions between samples and background variables were evaluated by corresponding two-way (sample by background variable) analyses of variance and analyses of covariance, with a separate analysis for each background variable (dichotomized, where necessary). These two-way analyses were carried out by multiple regression methods, each main effect being adjusted for the other main effect and the interaction being adjusted for all main effects. Interactions between samples and May SAT scores (dichotomized) were also evaluated by corresponding two-way (sample by May SAT score) analyses of variance and analyses of covariance. Interactions with May SAT scores were excluded in the analyses where the same May SAT score was also the dependent variable or covariate (e.g., the May SAT-verbal score was excluded in the analysis of variance of the May SAT-verbal score and in the analysis of covariance of the October SAT-verbal score).[3]

3. Background variables in continuous form and May SAT scores were dichotomized at their combined medians for the three samples; background variables in discrete form were dichotomized so that the two categories were meaningful and as equal in size as possible. The dichotomies for all variables are as follows: sex (male, female), ethnicity (white, all others), father's education (business or trade school or less, some college or more), mother's education (business or trade school or less, some college or more), parents' income ($26,999 or less, $27,000 or more), financial need (does not need aid, needs aid), high school type (public, other), high school program (academic, all others), high school rank (second fifth or less, top fifth), high school GPA (3.49 or less, 3.50 or more), educational aspiration (bachelor's degree or less, master's degree or more—other or undecided excluded), May SAT-verbal score (449 or less, 450 or more), and May SAT-mathematical score (499 or less, 500 or more).

Retest Reliability and Concurrent Validity. Sample differences in the product-moment correlations between corresponding SAT scores for May and October were appraised by a χ^2 test (Snedecor and Cochran 1967). Differences in the product-moment correlations of the SAT scores with the criterion variables were also evaluated by this χ^2 test.[4] Interactions between samples and background variables in both kinds of analyses were evaluated sequentially by the same χ^2 test:

1. An overall test was made of the correlations in the six subsamples formed by dividing each sample on the basis of a background variable (dichotomized the same way as in the analyses of variance and analyses of covariance). For instance, in the case of sex, the six subsamples were male control, male not encouraged, male encouraged, female control, female not encouraged, and female encouraged.

2. If this test was significant, follow-up tests were made of the correlations in the three subsamples at the same level of the background variable. In the case of sex, one level was male, and its subsamples were male control, male not encouraged, and male encouraged; the other level was female, and its subsamples were female control, female not encouraged, and female encouraged.

This process was carried out separately for each background variable. Interactions between samples and May SAT scores in the analyses of the correlations of the SAT scores with the criteria were evaluated in the same way. Interactions with high school rank, high school GPA, and May SAT scores were excluded in these analyses when the correlations were based on the same variables (for example, May SAT-verbal score was excluded in the analysis of the correlations of May SAT-verbal score with high school rank).

RESULTS AND DISCUSSION

Background Characteristics

The frequency distributions for the background variables in the three samples appear in Tables 1 through 11 along with the corresponding χ^2s and F ratios. The χ^2 for mother's education was significant ($p < .05$); none of the other χ^2s or F ratios was significant. The sample differences for mother's education were mainly due to a relatively small number of mothers with some college education in the control group and a comparatively small number with a high school diploma in the not encouraged group. The general similarity among the samples in these background variables indicates that self-selection in the examinees who returned for retesting did not have any differential

4. All correlations of SAT scores with high school rank have been reflected so that positive correlations with this criterion, like those with high school GPA, represent positive associations with academic performance.

Table 1. Sex of the Samples

Sex	Sample		
	Control (N = 1,248)	*Not Encouraged (N = 1,229)*	*Encouraged (N = 1,272)*
Male	632	593	613
Female	616	636	659

Note: $\chi^2(2) = 1.95, p > .05$.

Table 2. Ethnicity of the Samples

Ethnicity	Sample		
	Control (N = 1,158)	*Not Encouraged (N = 1,148)*	*Encouraged (N = 1,164)*
Black	40	45	49
Oriental	38	40	35
Puerto Rican	14	19	20
White	1,027	1,015	1,031
American Indian, Chicano, or other	39	29	29

Note: $\chi^2(8) = 4.49, p > .05$. Ethnicity was not ascertained for an additional 90 examinees in the control group, 81 in the not encouraged group, and 108 in the encouraged group.

Table 3. Father's Education of the Samples

Education	Sample		
	Control (N = 1,154)	*Not Encouraged (N = 1,137)*	*Encouraged (N = 1,156)*
Grade school	48	41	33
Some high school	63	81	67
High school diploma	233	231	243
Business or trade school	65	69	80
Some college	221	225	217
Bachelor's degree	204	177	214
Some graduate or professional school	63	63	55
Graduate or professional degree	257	250	247

Note: $\chi^2(14) = 11.92, p > .05$. Father's education was not ascertained for an additional 94 examinees in the control group, 92 in the not encouraged group, and 116 in the encouraged group.

Table 4. Mother's Education of the Samples

Education	Sample		
	Control (N = 1,151)	*Not Encouraged (N = 1,131)*	*Encouraged (N = 1,147)*
Grade school	35	32	24
Some high school	67	81	54
High school diploma	420	374	418
Business or trade school	97	77	108
Some college	198	229	221
Bachelor's degree	133	111	130
Some graduate or professional school	58	88	68
Graduate or professional degree	143	139	124

Note: $\chi^2(14) = 28.74, p < .05$. Mother's education was not ascertained for an additional 97 examinees in the control group, 98 in the not encouraged group, and 125 in the encouraged group.

Table 5. Parents' Income of the Samples

Income	Sample		
	Control (N = 1,010)	*Not Encouraged (N = 1,011)*	*Encouraged (N = 1,051)*
Less than $3,000	5	11	7
$ 3,000 - $ 5,999	14	17	16
$ 6,000 - $ 8,999	39	36	45
$ 9,000 - $11,999	52	47	48
$12,000 - $14,999	63	59	62
$15,000 - $17,999	75	90	79
$18,000 - $20,999	85	83	85
$21,000 - $23,999	83	84	88
$24,000 - $26,999	76	86	73
$27,000 - $29,999	86	77	104
$30,000 - $34,999	121	111	121
$35,000 - $39,999	77	82	89
$40,000 - $44,999	63	70	69
$45,000 - $49,999	36	42	35
$50,000 or more	135	116	130

Note: $\chi^2(28) = 14.45, p > .05$. Parents' income was not ascertained for an additional 238 examinees in the control group, 218 in the not encouraged group, and 221 in the encouraged group.

Table 6. Financial Need of the Samples

Need	Sample		
	Control (N = 1,139)	*Not Encouraged (N = 1,145)*	*Encouraged (N = 1,127)*
Needs aid	175	185	182
Does not need aid	964	929	945

Note: $\chi^2(2) = .66, p > .05$. Needs financial aid was not ascertained for an additional 109 examinees in the control group, 115 in the not encouraged group, and 145 in the encouraged group.

Table 7. High School Type of the Samples

	Sample		
Type	*Control (N = 1,071)*	*Not Encouraged (N = 1,042)*	*Encouraged (N = 1,062)*
Public	756	738	783
Other	315	304	279

Note: $\chi^2(2) = 3.20$, $p > .05$. High school type was not ascertained for an additional 177 examinees in the control group, 187 in the not encouraged group, and 210 in the encouraged group.

Table 8. High School Programs of the Samples

	Sample		
Program	*Control (N = 1,069)*	*Not Encouraged (N = 1,019)*	*Encouraged (N = 1,043)*
Academic	910	882	877
General	112	95	120
Career-oriented or other	47	42	46

Note: $\chi^2(4) = 2.85$, $p > .05$. High school program was not ascertained for an additional 179 examinees in the control group, 210 in the not encouraged group, and 229 in the encouraged group.

Table 9. High School Rank of the Samples

	Sample		
Rank	*Control (N = 990)*	*Not Encouraged (N = 965)*	*Encouraged (N = 997)*
Highest tenth	218	227	205
Second tenth	238	215	213
Second fifth	266	273	293
Middle fifth	236	227	260
Fourth fifth	32	20	26
Lowest fifth	0	3	0
Mean	26.95	26.50	27.91
Standard deviation	18.15	17.98	17.84

Note: $F(2,2949) = 1.59$, $p > .05$. High school rank was not ascertained for an additional 258 examinees in the control group, 264 in the not encouraged group, and 275 in the encouraged group.

Table 10. High School GPA of the Samples

	Sample		
GPA	*Control (N = 884)*	*Not Encouraged (N = 847)*	*Encouraged (N = 867)*
.50 - .99	0	1	1
1.00 - 1.49	1	1	2
1.50 - 1.99	12	12	17
2.00 - 2.49	72	62	62
2.50 - 2.99	192	173	183
3.00 - 3.49	268	277	291
3.50 - 3.99	219	220	230
4.00	120	101	81
Mean	3.28	3.28	3.25
Standard Deviation	.55	.55	.55

Note: $F(2,2595) = .70$, $p > .05$. High School GPA was not ascertained for an additional 364 examinees in the control group, 382 in the not encouraged group, and 405 in the encouraged group.

Table 11. Educational Aspiration of the Samples

	Sample		
Education	*Control (N = 1,060)*	*Not Encouraged (N = 1,032)*	*Encouraged (N = 1,039)*
Two-year specialized training	16	9	17
Associate's degree	17	10	21
Bachelor's degree	273	252	295
Master's degree	316	328	290
Doctorate or other professional degree	241	247	211
Other or undecided	197	186	205

Note: $\chi^2(10) = 16.23$, $p > .05$. Educational aspiration was not ascertained for an additional 188 examinees in the control group, 197 in the not encouraged group, and 233 in the encouraged group.

effect, at least as far as these diverse characteristics were concerned.

Score Level[5]

Analyses of Variance of Initial Scores. The means and standard deviations for the May SAT scores in the three samples appear in Table 12; the corresponding one-way analyses of variance are summarized in Table 13. The SAT-verbal score means were 449.05 for the control group, 451.97 for the not encouraged group, and 450.55 for the encouraged group; the SAT-mathematical score means were 501.05, 501.14, and 497.96, respectively. Neither of

5. Tables containing the means and standard deviations for the May, October, and covariance-adjusted October SAT scores in the subsamples defined by the background variables and May SAT scores, together with summaries of the corresponding two-way analyses of variance and analyses of covariance, are available from the author.

Table 12. Means and Standard Deviations (SD) for Initial, Retest, and Covariance-Adjusted Retest Scores

	Sample								
	Control			*Not Encouraged*			*Encouraged*		
Score	*N*	*Mean*	*SD*	*N*	*Mean*	*SD*	*N*	*Mean*	*SD*
SAT-verbal scores									
Initial	1,248	449.05	97.69	1,229	451.97	93.45	1,272	450.55	94.75
Retest	1,209	465.43	101.06	1,194	470.55	97.13	1,225	465.73	97.82
Adjusted retest	1,209	466.72	49.65	1,194	468.74	48.07	1,225	466.22	47.30
SAT-mathematical scores									
Initial	1,248	501.05	100.89	1,229	501.14	93.91	1,272	497.96	99.42
Retest	1,203	506.51	104.86	1,190	509.34	100.17	1,229	504.78	101.76
Adjusted retest	1,203	506.19	51.55	1,190	507.60	53.69	1,229	506.40	52.89

Table 13. Summary of One-way Analyses of Variance and Analyses of Covariance of Initial and Retest Scores

	SAT-verbal			*SAT-mathematical*		
Source	*df*	*Mean Square*	*F*	*df*	*Mean Square*	*F*
			Analysis of Variance of Initial Scores			
Sample	2	2,645.76	.29	2	4,119.85	.43
Error	3,746	9,093.54		3,746	9,641.26	
			Analysis of Variance of Retest Scores			
Sample	2	9,935.34	1.02	2	6,394.47	.61
Error	3,625	9,747.81		3,619	10,471.37	
			Analysis of Covariance of Retest Scores[a]			
Sample	2	2,145.83	.92	2	654.23	.24
Error	3,624	2,339.96		3,618	2,784.39	

Note: None of the F ratios are significant ($p > .05$).

a. F (2, 3622) = .25 ($p > .05$) for the within-group regression coefficients in the SAT-verbal score analysis, and F (2,3616) = 1.44 ($p > .05$) for those in the SAT-mathematical score analysis.

the F ratios for the corresponding one-way analyses of variance was significant ($p > .05$). And none of the F ratios for the interactions with the background variables and May SAT scores in the two-way analyses of variance was significant ($p > .05$).

These results complement those for the background variables in indicating that the samples of retested examinees were comparable in their initial performance, too. This point is reinforced by the analyses of interactions, which established that the similarity of the samples extended to a variety of subgroups.

Analyses of Variance of Retest Scores. The means and standard deviations for the October SAT scores in the three samples are shown in Table 12; the analyses of variance are summarized in Table 13. The SAT-verbal score means were 465.43 for the control group, 470.55 for the not encouraged group, and 465.73 for the encouraged group; the SAT-mathematical score means were 506.51, 509.34, and 504.78, respectively. The F ratios for the one-way analyses of variance were not significant ($p > .05$). Similarly, the F ratios for the interactions with the background variables and May SAT scores in the two-way analyses of variance were not significant ($p > .05$).

These consistently negative results strikingly demonstrate that the samples did not differ in their retest scores, even when various subgroups were examined. The present findings, taken together with the uniformly negative results in the analyses of initial scores, imply that access to the disclosed material and the motivation provided by the encouraging letter did not affect the level of retest performance, either for the total samples or the subsamples.

Analyses of Covariance of Retest Scores. The covariance-adjusted means and standard deviations for the October SAT scores in the three samples are reported in Table 12; the analyses of covariance are summarized in Table 13. The SAT-verbal score adjusted means were 466.72 for the control group, 468.74 for the not encouraged group, and 466.22 for the encouraged group; the SAT-mathematical score adjusted means were 506.19, 507.60, and 506.40, respectively. The F ratios for the one-way analyses of covariance were not significant ($p > .05$). In addition, the F ratios for the interactions with background variables and May SAT scores in the two-way analyses of covariance were not significant ($p > .05$).[6]

These findings are congruent with the preceding results for the October SAT scores in demonstrating that the samples and subsamples did not differ in their retest scores and in suggesting that the disclosed material and the motivating letter uniformly failed to have an impact on the level of retest performance. The close resemblance between the two sets of results is not surprising, even though the present analyses take into account initial differences in the samples and the other analyses do not, for the samples were observed to be similar in the analyses of May SAT scores.

Retest Reliability[7]

The correlations between the corresponding May and October SAT scores in the three samples appear in Table 14 together with the χ^2s. The SAT-verbal correlations were .87 for the control group, .87 for the not encouraged group, and .88 for the encouraged group; the corresponding SAT-mathematical correlations were .87, .84, and .85. The χ^2 was significant ($p < .05$) for SAT-mathematical scores but not for SAT-verbal scores.

In the analyses of the SAT-verbal correlations in the subsamples defined by the background variables and May SAT scores, the overall χ^2s were significant ($p < .05$) for high school rank, educational aspiration, and May SAT-mathematical score subsamples. However, in the follow-up analyses of the correlations in the subsamples at each level of these three variables, none of the χ^2s was significant ($p > .05$). In the parallel analyses of the SAT-mathematical correlations, the overall χ^2s were significant ($p < .05$) for sex, ethnicity, and high school rank subsamples. In the follow-up analyses, the corresponding χ^2s were significant ($p < .05$) for one level of sex (male), ethnicity (nonwhite), and high school rank (top fifth of class). For the male subsamples, the correlations were .88 for the control group, .84 for the not encouraged group, and .86 for the encouraged group; for the nonwhite subsamples, .90, .84, and .91; and for the top fifth of the class subsamples, .85, .80 and .85. These sets of correlations resemble those for the total samples, with slightly lower correlations occurring for the not encouraged group than the other groups.

These results indicate that the sample differences in retest reliability were very minor, being limited to extremely small divergences for SAT-mathematical scores. The subsample findings also suggest that these differences were not uniform throughout the samples but stemmed from a few isolated subgroups of examinees. Whether or not this outcome is traceable to variations in sample composition or in retest performance cannot be determined. In any event, it appears that the disclosed material and the letters had no more than a negligible impact on score stability for the samples as a whole as well as the various subgroups.

Concurrent Validity[8]

Initial Scores. The correlations of the May SAT scores with high school rank and high school GPA in the three samples are reported in Table 15 together with the χ^2s. The SAT-verbal score correlations with high school rank were .51 for the control group, .42 for the not encouraged group, and .46 for the encouraged group; the corresponding SAT-mathematical score correlations were .53, .49, and .50. The χ^2 was significant ($p < .05$) for SAT-verbal scores but not for SAT-mathematical scores. The SAT-verbal correlations with high school GPA were .51, .46, and .48 in the three samples; the SAT-mathematical correlations were .51, .47, and .49, respectively. The χ^2s for SAT-verbal and SAT-mathematical scores were not significant ($p > .05$).

In the analyses of the SAT-verbal correlations with high school rank in the subsamples defined by the background variables and May SAT scores, the overall χ^2s were significant ($p < .05$) for sex, ethnicity, mother's education, parents' income, financial need, high school type, high school program, and May SAT-mathematical score subsamples. In the follow-up analyses of the correlations in the subsamples at each level of these eight variables, the χ^2s were significant ($p < .05$) for one level of sex (males), mother's education (some college or more), financial need (needs aid), high school type (public school), high school program (academic program), and SAT-mathematical score (499 or less). For the male subsamples, the correlations were .51 for the control group, .35 for the not encouraged group, and .43 for the encouraged group; for the mothers with some college or more subsamples, .54, .38, and .50; for the needs aid subsamples, .50, .40, and .45; for the

6. The F ratio for the within-group regression coefficients was significant ($p < .05$) in the analysis of SAT-mathematical score with high school rank, and hence its covariance results are uninterpretable. None of the F ratios for the other analyses were significant ($p > .05$).

7. Tables containing the correlations between corresponding May and October SAT scores in the subsamples defined by the background variables, together with the corresponding χ^2s, are available from the author.

8. Tables containing the correlations of the May and October SAT scores with high school rank and high school GPA in the subsamples defined by the background variables, together with the corresponding χ^2s, are available from the author.

Table 14. Correlations Between Initial and Retest Scores

	Sample						
	Control		Not Encouraged		Encouraged		
Score	N	r	N	r	N	r	χ^2
SAT-verbal scores	1,209	.87	1,194	.87	1,225	.88	.46
SAT-mathematical scores	1,203	.87	1,190	.84	1,229	.85	6.35*

Note: All the correlations are significant ($p < .01$).
*$p < .05$

Table 15. Correlations of Initial and Retest Scores with Criteria

	Sample						
	Control		Not Encouraged		Encouraged		
Score	N	r	N	r	N	r	χ^2
High School Rank							
SAT-verbal scores							
Initial	990	.51	965	.42	997	.46	7.79*
Retest	959	.52	941	.43	965	.47	6.26*
SAT-mathematical scores							
Initial	990	.53	965	.49	997	.50	1.21
Retest	955	.54	939	.49	969	.52	1.71
High School GPA							
SAT-verbal scores							
Initial	884	.51	847	.46	867	.48	1.80
Retest	860	.51	828	.48	840	.49	.84
SAT-mathematical scores							
Initial	884	.51	847	.47	867	.49	1.01
Retest	856	.53	826	.49	843	.51	1.17

Note: All the correlations are significant ($p < .01$).
*$p < .05$

public school subsamples, .57, .44, and .48; for the academic program subsamples, .48, .37, and .45; and for the SAT-mathematical scores of 499 or less subsamples, .36, .18, and .26. These patterns of correlations are similar to those for the total samples, with the highest correlations for the control group and the lowest for the not encouraged group.

In the parallel analyses of the SAT-mathematical score correlations with this criterion, the overall χ^2s were significant ($p < .05$) for mother's education, financial need, high school type, high school program, and May SAT-verbal score subsamples. In the follow-up analyses, the χ^2 was significant ($p < .05$) for one level of high school program (nonacademic program). The correlations for the nonacademic program subsamples were .45 for the control group, .54 for the not encouraged group, and .29 for the encouraged group. These correlations diverge from those for the total samples, where the highest correlations occurred for the control group, not the not encouraged group.

In the analyses of the SAT-verbal and SAT-mathematical score correlations with high school GPA in the subsamples defined by the background variables and May SAT scores, none of the overall χ^2s was significant ($p > .05$).

The sample differences in concurrent validity, though largely restricted to the validity of SAT-verbal scores against high school rank in a scattered assortment of subgroups, are notable. They imply, in contrast to the previous findings in the analyses of background variables and May SAT scores, that the samples differed in their initial test performance, but in some aspect other than sheer level of performance.

Retest Scores. The correlations of the October SAT scores with high school rank and high school GPA in the three samples are given in Table 15 along with the χ^2s. The SAT-verbal correlations with high school rank were .52 for the control group, .43 for the not encouraged group, and .47

for the encouraged group; the SAT-mathematical correlations were .54, .49, and .52. The χ^2s were significant ($p < .05$) for SAT-verbal scores but not for SAT-mathematical scores. The corresponding correlations with high school GPA were .51, .48, and .49 for SAT-verbal scores; and .53, .49, and .51 for SAT-mathematical scores. The χ^2s for SAT-verbal and SAT-mathematical scores were not significant ($p > .05$).

In the analyses of the SAT-verbal score correlations with high school rank in the subsamples defined by the background variables and May SAT scores, the overall χ^2s were significant ($p < .05$) for sex, mother's education, parents' income, and high school type subsamples. In the follow-up analyses of the correlations in the subsamples at each level of these four variables, the χ^2s were significant ($p < .05$) for one level of sex (males), mother's education (some college or more), and high school type (public school). For the male subsamples, the correlations were .53 for the control group, .39 for the not encouraged group, and .42 for the encouraged group; for the mothers with some college or more subsamples, .54, .39, and .52; and for the public school subsamples, .57, .44, and .49. These patterns of correlations are similar to those for the total samples, with the highest correlations for the control group and the lowest for the not encouraged group.

In the parallel analyses of the SAT-mathematical score correlations with this criterion, the overall χ^2s were significant ($p < .05$) for ethnicity, mother's education, high school type, and May SAT-verbal score subsamples. In the follow-up analyses, none of the χ^2s was significant ($p > .05$).

In the analyses of the SAT-verbal score correlations with high school GPA in the subsamples defined by the background variables and May SAT scores, the overall χ^2 was significant ($p < .05$) for May SAT-verbal score subsamples. In the follow-up analyses, the χ^2 was significant ($p < .05$) for one level (449 or less). The correlations for the SAT-verbal score of 449 or less subsamples were .30 for the control group, .36 for the not encouraged group, and .18 for the encouraged group. These correlations deviate from the correlations for the total samples, which were all at about the same level. In the analyses of the SAT-mathematical score correlations with this criterion, none of the overall χ^2s was significant ($p > .05$).

The similarity between these findings and those for the May SAT scores is pronounced, with sample differences largely being limited to the SAT-verbal score correlations with high school rank, and some of the same subsamples being involved. The correspondence between the sample differences in concurrent validity in the two analyses suggests that the present results were due to variations in the composition of the samples, not variations in their retest performance. The disclosed material and the letters seem to have had little or no effect on the concurrent validity of the retest scores, either for the total samples or the subgroups within them.

CONCLUSIONS

The main conclusion of this study is that access to the disclosed test material had no appreciable effects on subsequent retest performance–whether or not that performance was defined in terms of the level, stability, or concurrent validity of the new scores. It also appears, though the evidence on this point is suggestive, that use of the material had no discernible effect either.

It should be emphasized that this investigation was primarily concerned with access to the disclosed material and only secondarily with its use. The design of this study guaranteed that the examinees in the experimental groups had the material, but did not ensure its use, even in the group that was encouraged to do so. The intrinsic motivation to use the disclosed material was probably low, judging from the extremely small number of examinees in the control group who requested the disclosed material. However, the absence of effects for the examinees most apt to use the material–those who were sent the encouraging letter, and those in the middle class and other subgroups that are highly motivated to attend very selective colleges and universities–implies that use had no impact.

A definitive answer about the effects of use on performance would be of some interest, but such an undertaking is fraught with conceptual and methodological difficulties. Inquiring about use in a study like the current one is not only prone to inaccuracy in the reports but also to incomplete and potentially unrepresentative data, for many examinees may simply fail to respond. Teaching the disclosed material in a classroom setting, along the lines of the Hale *et al.* (1980) study, produces results of limited generalizability to real life, where use is voluntary. And simply comparing the performance of examinees who request the disclosed material with those who do not inextricably confounds variables linked to requesting the material with variables connected to using it.

The general failure to find any effects in this study casts doubt on the line of reasoning that suggested retest performance might be altered. This reasoning rests on two propositions: (a) greater familiarity with the nature of a test, reduced test anxiety, and intensive drilling on the test's items enhance performance; and (b) these factors are influenced by using the disclosed material. It is entirely possible that the second proposition is incorrect, at least in the present context, because of the nature of the examinees and the test involved. First, the examinees may have been maximally familiar with the test and minimally anxious about it by the time they had received the disclosed material. All had been tested with the SAT in the May administration and routinely received *Taking the SAT* (Educational Testing Service 1979), an orientation booklet that contained a sample form of the SAT, when they registered for that administration. Furthermore, many of the examinees had undoubtedly been tested with the Preliminary Scholastic Aptitude Test and other tests that are similar to the SAT

during their school career, and at least some of them may have had access to previously disclosed forms of the SAT, especially the four that appear in a widely circulated publication, *4 SATs* (Educational Testing Service 1980). Hence, the familiarization and anxiety reduction provided by using the disclosed material may have already been accomplished. This speculation is consistent with the finding that practice on a test and other kinds of exposure to it have the greatest effect on score level for naïve examinees and that the gains diminish with repeated practice and exposure (See Bond 1981; Messick and Jungeblut 1981). Second, the SAT may not be influenced by drilling because the test includes few, if any, of the kinds of items on which performance can be improved by such practice. Systematic efforts are made to eliminate such items from the SAT (Donlon and Angoff 1971).

The negative results necessarily raise questions about the efficacy of the experimental operations, particularly the encouraging letter, the variables used to form the subgroups, and the criteria. The effectiveness of the letter, which is modeled after one employed in a study of test familiarization involving the Graduate Record Examinations Aptitude Test, is suggested by its marked effects, in that investigation, on the amount of time that the examinees used the material and on their test scores (Powers and Swinton 1982). The measures used to define the subgroups in the present investigation, though adequate for exploratory purposes, are not ideal. The May SAT scores are reasonable indexes of the ability to take advantage of the disclosed material. The background variables are adequate measures of key demographic characteristics but no more than substitutes for direct assessments of anxiety, motivation, familiarity with tests, and so forth. The value of the self-reported criteria, high school rank and high school GPA, is demonstrated by (a) their substantial correlations with May and October SAT scores in this study, (b) the accuracy of self-reported grades (see Baird 1976), and (c) the appreciable correlations between self-reported high school grades and recorded college grades (see Baird 1976).

REFERENCES

Anastasi, A. 1981. Diverse Effects of Training on Tests of Academic Intelligence. In *Issues in Testing: Coaching, Disclosure, and Ethnic Bias,* B. F. Green, ed. San Francisco: Jossey-Bass.

Baird, L. L. 1976. *Using Self-reports to Predict Student Performance.* New York: College Entrance Examination Board.

Bond, L. 1981. Bias in Mental Tests. In *Issues in Testing: Coaching, Disclosure, and Ethnic Bias,* B. F. Green, ed. San Francisco: Jossey-Bass.

Brown, R. 1980. *Searching for the Truth About "Truth in Testing" Legislation.* Denver, Colo.: Education Commission of the States.

Donlon, T. F., and Angoff, W. H. 1971. The Scholastic Aptitude Test. In *The College Board Admissions Testing Program: A Technical Report on Research and Development Activities Relating to the Scholastic Aptitude Test and Achievement Tests,* W. H. Angoff, ed. New York: College Entrance Examination Board.

Educational Testing Service 1979. *Taking the SAT.* New York: College Entrance Examination Board.

Educational Testing Service 1980. *4 SATs.* New York: College Entrance Examination Board.

Educational Testing Service 1981a. *ATP Guide for High Schools and Colleges 1981-82.* New York: College Entrance Examination Board.

Educational Testing Service 1981b. *National Report, College-bound Seniors, 1981.* New York: College Entrance Examination Board.

Hale, G. A., *et al.* 1980. *Effects of Item Disclosure on TOEFL Performance* (TOEFL RR 8). Princeton, N.J.: Educational Testing Service.

Linn, R. L. 1982. Admissions Testing on Trial. *American Psychologist* 37: 279-291.

Messick, S. 1980. *The Effectiveness of Coaching for the SAT: Review and Reanalysis of Research from the Fifties to the FTC.* Princeton, N.J.: Educational Testing Service.

Messick, S. 1981. The Controversy Over Coaching: Issues of Effectiveness and Equity. In *Issues in Testing: Coaching, Disclosure, and Ethnic Bias,* B. F. Green, ed. San Francisco: Jossey-Bass.

Messick, S., and Jungeblut, A. 1981. Time and Method in Coaching for the SAT. *Psychological Bulletin* 89: 191-216.

Olsen, M., and Schrader, W. B. 1959. *The Use of Preliminary and Final Scholastic Aptitude Test Scores in Predicting College Grades* (ETS SR 59-19). Princeton, N.J.: Educational Testing Service.

Powers, D. E., and Swinton, S. S. 1982. *Effects of Self-study of Test Familiarization Materials for the Analytical Section of the GRE Aptitude Test* (GREB RR 79-9R). Princeton, N.J.: Educational Testing Service.

Snedecor, G. W., and Cochran, W. G. 1967. *Statistical Methods.* Sixth Ed. Ames, Iowa: Iowa State University.

Strenio, A. Jr. 1979. *The Debate Over Open Versus Secure Testing: A Critical Review.* National Consortium on Testing Staff Circular No. 6. Cambridge, Mass.: Huron Institute.

April 6, 1981, Test-takers May Ask for and Get Answers to SATs Next Year, College Board Decides. *Chronicle of Higher Education,* p.1.